Capstone Project - The Battle of the Neighborhoods (Week 2)¶

Applied Data Science Capstone by IBM/Coursera¶

Table of contents¶

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Introduction: Business Problem ¶

A famous chef of Italian Cuisine wants to open a 5 stars restaurant in Bristol.

He is not interested in a commercial area, not excluded, but mostly to a very selected and quiet one. The presence of the competitors of same cuisine should be low but, he is counting of support of other international cuisine for create highly gourmet event.

The focus is to find a suitable neighborhood for the purpose.

The background¶

Bristol is a big city in South West England and has 35 wards and about 100 neighborhoods. It's a very huge district but only the urban will be considered.

Data ¶

Collecting it¶

In order to gathering the right data, different sources will be used which will require techniques of sgrabbing.

The data related of:

- *Wards and Population* will be exported in csv format from https://www.doogal.co.uk/AdministrativeAreas.php?district=E06000023
- *Neighborhood* will be sgrabbed from wikipedia pages and joing with wikimedia API will fetch the coordinates
- *Crime* data will downloaded from https://opendata.bristol.gov.uk/explore/dataset/crime-recorded-by-police-by-selected-offence-groups-in-bristol-by-ward/export/?disjunctive.ward_name
- *House price* will be get from https://www.bristolpost.co.uk/news/bristol-news/house-prices-bristol-2018-postcode-2279067

The data to collect for each area for restaurants will be provided by using the Foursquare APIs where for each neighborhood coordinates will be used the explore functionality within 1000 meter radius and limited to 100 venues and the price parameter with value 3,4

How it will be used to solve the problem¶

The main purpose of data is to show plots with Folium package by using the coordinates which suggest the right area. Additionally techniques of clustering will be used for aggregate those neighborhoods with strong similarities.

Wards¶

The *Doogal.co.uk* provides the all the available post codes and several information like:

- District
- Wards
- Lsoa
- Population
- Average Incoming
- Index of Multiple Deprivation (see: https://www.bristol.gov.uk/documents/20182/32951/Deprivation+in+Bristol+2019.pdf/ff3e5492-9849-6300-b227-1bdff2779f80)

The limit of this dataset is to have just the Ward information.

The dataset will filtered by the following

- **'District' == 'Bristol, City of'** which the column related to district of Bristol City
- **'Rural/urban' == 'Urban city and town'** which the column for specifying if the record

Ref: https://www.doogal.co.uk/AdministrativeAreas.php?district=E06000023

The dataset has the following details:

```
Shape:  (26717, 47)

Columns:
['Postcode', 'In Use?', 'Latitude', 'Longitude', 'Easting', 'Northing',
       'Grid Ref', 'County', 'District', 'Ward', 'District Code', 'Ward Code',
       'Country', 'County Code', 'Constituency', 'Introduced', 'Terminated',
       'Parish', 'National Park', 'Population', 'Households', 'Built up area',
       'Built up sub-division', 'Lower layer super output area', 'Rural/urban',
       'Region', 'Altitude', 'London zone', 'LSOA Code', 'Local authority',
       'MSOA Code', 'Middle layer super output area', 'Parish Code',
       'Census output area', 'Constituency Code',
       'Index of Multiple Deprivation', 'Quality', 'User Type', 'Last updated',
       'Nearest station', 'Distance to station', 'Postcode area',
       'Postcode district', 'Police force', 'Water company', 'Plus Code',
       'Average Income']

The unique values of Districts:
['Bristol, City of' 'South Gloucestershire' 'North Somerset'
 'Bath and North East Somerset' 'Sedgemoor' 'Mendip']

The unique values of Rural/urban:
['Urban city and town' 'Rural town and fringe' 'Rural village'
 'Rural hamlet and isolated dwellings']
```

Filter by District equals to Bristol, City and Urban city.

```
Current shape :  (11542, 47)

The unique values of Ward:
['Lawrence Hill' 'Central' 'Ashley' 'Cotham' 'Hotwells and Harbourside'
 'Horfield' 'Southmead' 'Westbury-on-Trym and Henleaze'
```

```
 'Henbury and Brentry' 'Avonmouth and Lawrence Weston'
 'Hartcliffe and Withywood' 'Hengrove and Whitchurch Park' 'Bedminster'
 'Bishopsworth' 'Stockwood' 'Knowle' 'St George Central' 'Hillfields'
 'Eastville' 'St George Troopers Hill' 'Frome Vale' 'Lockleaze'
 'Southville' 'Windmill Hill' 'Filwood' 'Brislington West'
 'Brislington East' 'Easton' 'Clifton Down' 'St George West' 'Redland'
 'Bishopston and Ashley Down' 'Clifton' 'Stoke Bishop']

The unique values of Districts:
['Bristol, City of']

The unique values of Rural/urban:
['Urban city and town']
```

Out[5]:

	Postcode	Grid Ref	County	District	Ward	District Code	Ward Code	County Code	Constituency	Population	...	Lower layer super output area	Reg
0	BS1 1AD	ST602734	Bristol	Bristol, City of	Lawrence Hill	E06000023	E05010907	E11000003	Bristol West	NaN	...	Bristol 054D	Sou Wes
1	BS1 1BU	ST588729	Bristol	Bristol, City of	Central	E06000023	E05010892	E11000003	Bristol West	15.0	...	Bristol 032B	Sou Wes
2	BS1 1DA	ST587730	Bristol	Bristol, City of	Central	E06000023	E05010892	E11000003	Bristol West	NaN	...	Bristol 032B	Sou Wes
3	BS1 1DB	ST587730	Bristol	Bristol, City of	Central	E06000023	E05010892	E11000003	Bristol West	NaN	...	Bristol 032B	Sou Wes
4	BS1 1DD	ST601735	Bristol	Bristol, City of	Lawrence Hill	E06000023	E05010907	E11000003	Bristol West	NaN	...	Bristol 056A	Sou Wes

5 rows Ã— 21 columns

fetching map shapes of Wards¶

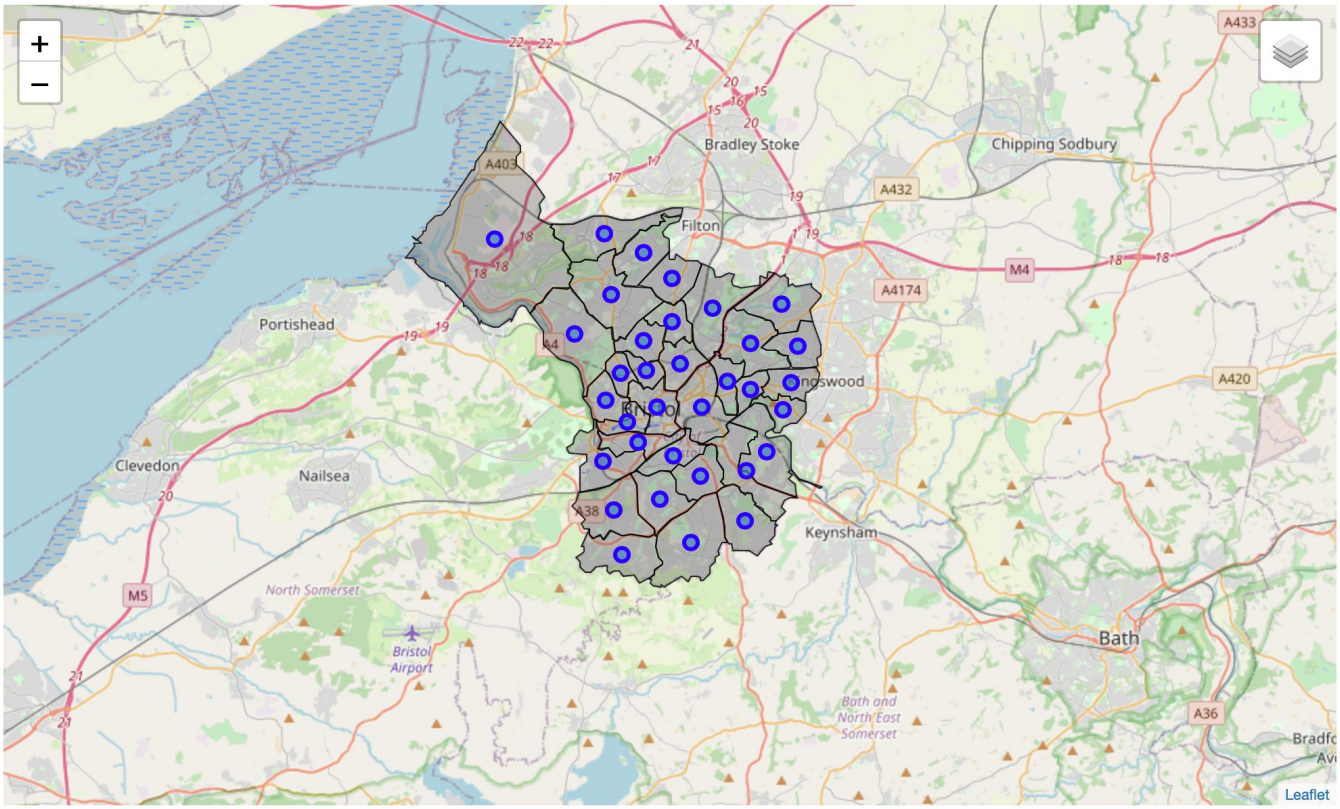
The wards are available through the opendata service of Bristol concil; the same data contains the the **polygons** and coordinates of **map shapes** which we are interest of because it's missing inside the previous dataset

references:

- https://opendata.bristol.gov.uk/explore/dataset/wards/table/?location=11,51.47092,-2.61607&basemap=jawg.streets
- https://opendata.bristol.gov.uk/explore/dataset/bristol-ward-areas/information/?disjunctive.2016_ward_name

	OBJECTID	Name	No of Councillors	Ward ID	geo_shape	geo_poi
0	3	Henbury & Brentry	2	E05010901	{"type": "Polygon", "coordinates": [[[-2.64716...	51.5089723054,-2
1	25	Lawrence Hill	2	E05010907	{"type": "Polygon", "coordinates": [[[-2.58406...	51.4545915255,-2
2	33	St George Troopers Hill	1	E05010911	{"type": "Polygon", "coordinates": [[[-2.54152...	51.4534343901,-2
3	11	Eastville	2	E05010897	{"type": "Polygon", "coordinates": [[[-2.53282...	51.4747087486,-2

	OBJECTID	Name	No of Councillors	Ward ID	geo_shape	geo_poi
4	15	Southville	2	E05010914	{"type": "Polygon", "coordinates": [[[-2.61227...	51.4433202033,-2



Fecth Lsoa¶

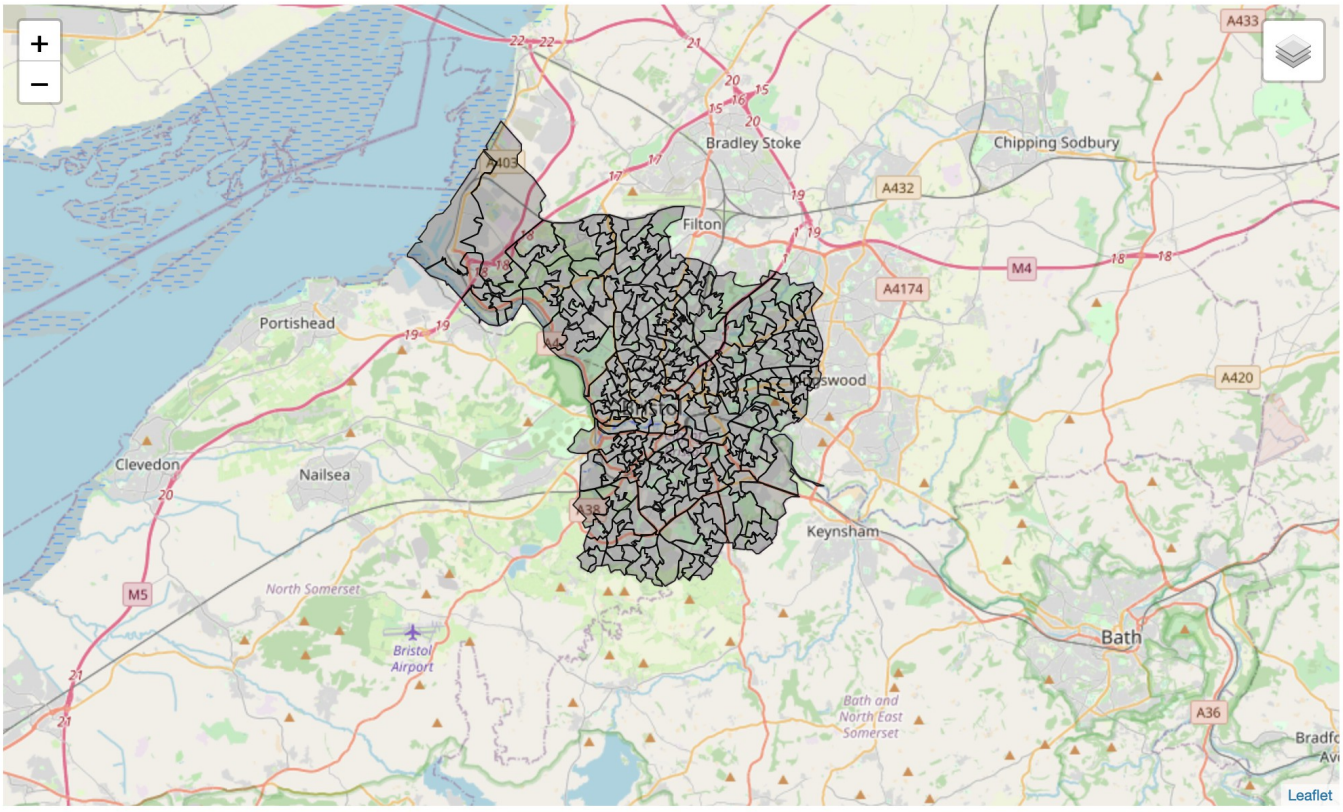
The Lsoa areas are a deeper subdivision of wards and we are going to map our neighbourhoods with it.

Ref. https://data.opendatasoft.com/explore/dataset/lsoa110%40bristol/map/?location=11,51.47732,-2.58659&basemap=jawg.sunny

	OBJECTID	LSOA11 Code	LSOA11 Name	LSOA11 Local name	MSOA11 Code	Ward code	Area square mteres	Perimeter (m)	MI_PRINX	
0	14	E01014730	Bristol 037A	Lower Totterdown	E02003048	E05002006	280859.28125	3532.973145	892	{"type": "coordin 2.57611
1	3	E01014689	Bristol 002B	Southmead Central	E02003013	E05002000	266356.62500	4146.445312	1225	{"type": "coordin 2.59576
2	26	E01014518	Bristol 046A	Ilchester Crescent	E02003057	E05001976	536719.50000	4171.108398	531	{"type": "coordin 2.61127
3	30	E01014596	Bristol 051C	Fair Furlong	E02003062	E05001987	310003.46875	3115.560547	606	{"type": "coordin 2.60607
4	64	E01014537	Bristol 035F	Eldonwall	E02003046	E05001978	346547.50000	3192.706787	668	{"type": "coordin 2.54735
...	...	...	...	...	...	...	...	...	...	...

	OBJECTID	LSOA11 Code	LSOA11 Name	LSOA11 Local name	MSOA11 Code	Ward code	Area square mteres	Perimeter (m)	MI_PRINX	g
258	232	E01014524	Bristol 038B	Birchwood Road	E02003049	E05001977	314044.31250	3070.859619	773	{"type": "coordin 2.53906
259	221	E01014603	Bristol 001C	Brentry East	E02003012	E05001988	504946.03125	3934.161865	1222	{"type": "coordin 2.60631
260	229	E01014704	Bristol 049C	Woodleigh Gardens	E02003060	E05002002	394265.84375	3304.940918	380	{"type": "coordin 2.55330
261	249	E01014515	Bristol 050A	Highridge Common	E02003061	E05001976	321753.43750	3514.458008	49	{"type": "coordin 2.61952
262	245	E01014650	Bristol 044E	Broadwalk	E02003055	E05001994	242799.21875	3229.918701	643	{"type": "coordin 2.58024

The size of dataset is: 263 rows and 13 columns



Neighbourhoods of Bristol¶

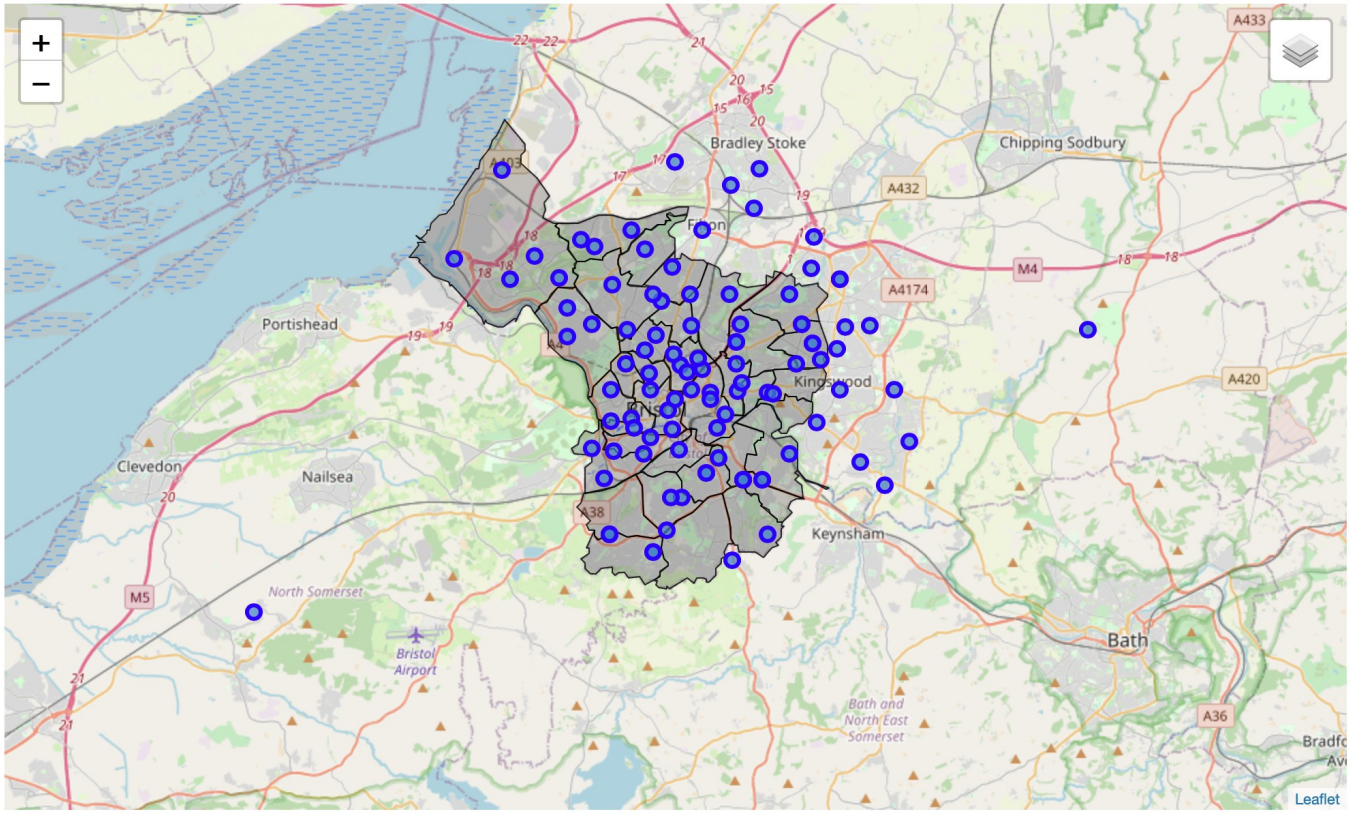
The data has been fetch by sgrabbing wikipedia pages and combined to Wikimedia API for getting the coordinates.

Missing coordinates have been manually fetched from internet as well as the mapping between ward and neighbourhoods.

The sgrab starts from: `https://en.wikipedia.org/wiki/Cabot,_Bristol#Statistics`

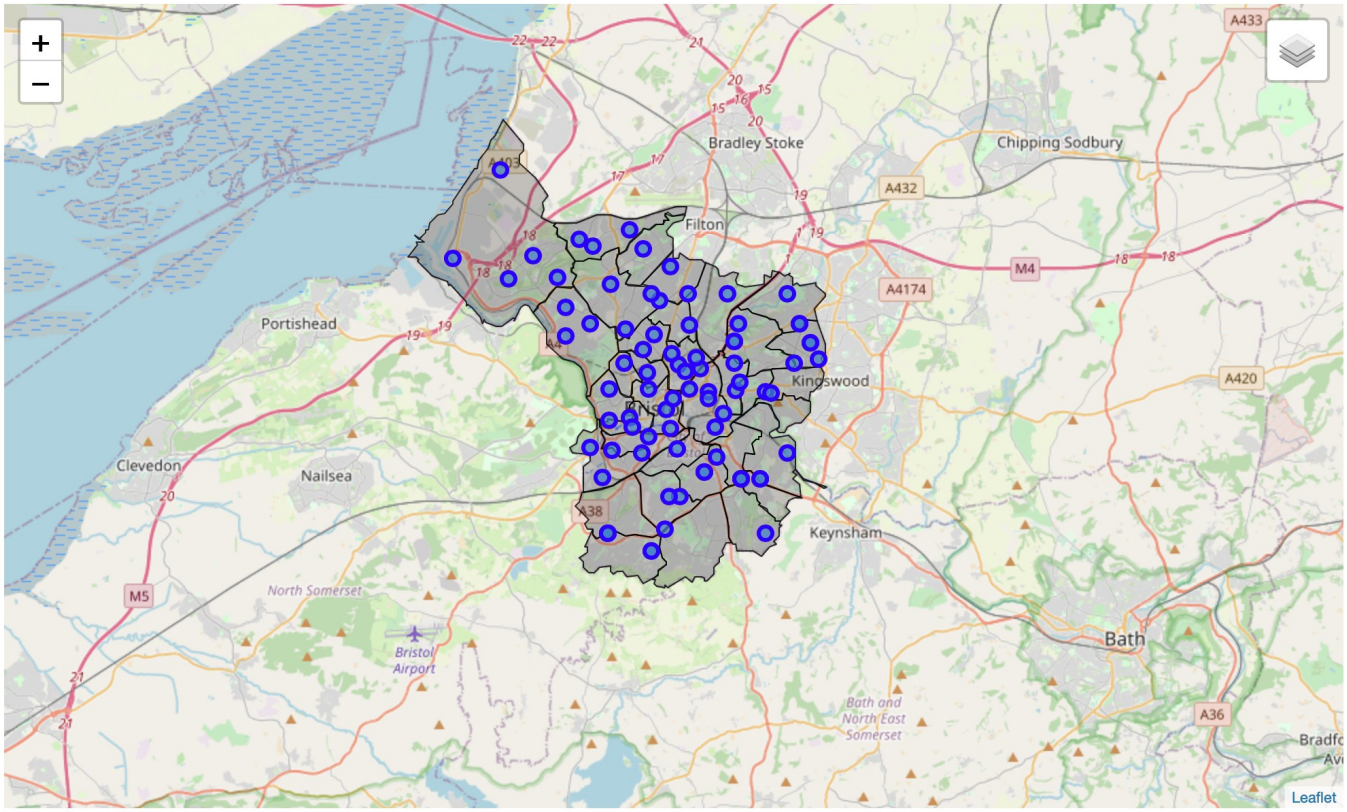
The API: https://en.wikipedia.org/w/api.php?action=query&prop=coordinates&titles={}&format=json

```
The dataset has:

Shape (102, 3)

First five records:
           name   latitude  longitude
0  City centre  51.453632  -2.591341
1   Arnos Vale  51.431600  -2.553600
2  Ashley Down  51.480000  -2.580000
3  Ashton Gate  51.440600  -2.619050
4  Ashton Vale  51.432193  -2.623522

Columns:
name          object
latitude     float64
longitude    float64
```



Because it's not possible to proper mapping the neighbourhoods and the wards, firstly we are going to see if there are visually coordinates which are not belonging to the ward we are interested of.

The mapping between neighbourhoods and the wards of City of Bristol is performed by evaluating if the neighbourhoods coordinates fell inside a ward polygon.




```
Neighbourhoods dataset size  (81, 9)
```

Fetching the house price¶

The following step will add details to the dataset related to average house price. The minimum details of price is based at first number of second part of Postcode ex: BS1 1. The information is fetched from the website which is providing interesting statistics about the cities in United Kingdom: housing market, population, income, unemployment

The website *plumplot.co.uk* will be sgrabbed.

Ref: https://www.plumplot.co.uk/Bristol-house-prices.html

	Postcode2	Avg. House Price Â£k
0	BS1 1	187
1	BS1 2	226
2	BS1 3	247
3	BS1 4	256
4	BS1 5	386

Fetching crime details¶

Fetching information about the crimes in Bristol it's available from *opendata* of *Bistol Concil* and it's available only for Ward.

Additionally, the original dataset give the coordinates of Wards.

ref: https://opendata.bristol.gov.uk/explore/dataset/crime-recorded-by-police-by-selected-offence-groups-in-bristol-by-ward/export/?disjunctive.ward_name

	Ward Code	Ward Name	Ward Population	Ward Crimes	Ward Latitude	Ward Longitude
0	E05010887	Bedminster	12517	1036	51.437160	-2.621872
1	E05010901	Henbury & Brentry	12473	1150	51.508974	-2.621035

	Ward Code	Ward Name	Ward Population	Ward Crimes	Ward Latitude	Ward Longitude
2	E05010894	Clifton Down	11639	845	51.464970	-2.612657
3	E05010909	Redland	13088	772	51.475245	-2.601158
4	E05010912	St George West	6518	669	51.460117	-2.546940

Merging dataset¶

After collecting different datasets now the process continues with merging the census (Average Incoming Index of Multiple Deprivation) and house price

	Ward	Ward Code	Average Income	Index of Multiple Deprivation	Postcode2	Avg. House Price £k
0	Lawrence Hill	E05010907	39100	1501	BS1 1	187
1	Central	E05010892	47500	12154	BS1 1	187
2	Central	E05010892	47500	12154	BS1 1	187
3	Central	E05010892	47500	12154	BS1 1	187
4	Lawrence Hill	E05010907	29100	1881	BS1 1	187

Moving forward to the final ward dataset

	Ward Code	Ward Name	Ward Latitude	Ward Longitude	Average Income	Index of Multiple Deprivation	A
0	E05010885	Ashley	51.468310	-2.582673	44400.000000	9698.176600	34
1	E05010886	Avonmouth & Lawrence Weston	51.507334	-2.676404	38226.916803	8580.047308	24
2	E05010887	Bedminster	51.437160	-2.621872	43119.875776	15573.475155	31
3	E05010888	Bishopston & Ashley Down	51.481453	-2.586463	56094.258373	24239.947368	40
4	E05010889	Bishopsworth	51.421788	-2.616456	37048.484848	14629.829545	23
5	E05010890	Brislington East	51.440284	-2.538654	40835.074627	12991.216418	24
6	E05010891	Brislington West	51.434394	-2.549070	43799.619772	13685.300380	26
7	E05010892	Central	51.454415	-2.594336	45808.049113	12850.009550	30
8	E05010893	Clifton	51.456516	-2.619963	50983.238636	23870.107955	53
9	E05010894	Clifton Down	51.464970	-2.612657	56292.763158	25230.625000	45
10	E05010895	Cotham	51.465963	-2.599911	52680.487805	21720.739837	42
11	E05010896	Easton	51.462372	-2.558506	37942.013889	8262.618056	25
12	E05010897	Eastville	51.474710	-2.546927	41384.023669	9685.763314	27
13	E05010898	Filwood	51.425419	-2.592803	34159.854015	2816.689781	22
14	E05010899	Frome Vale	51.487091	-2.531199	37993.548387	10998.900293	30
15	E05010900	Hartcliffe & Withywood	51.407704	-2.612218	30537.915743	1579.166297	20
16	E05010901	Henbury & Brentry	51.508974	-2.621035	38743.076923	8907.323077	25
17	E05010902	Hengrove & Whitchurch Park	51.411542	-2.577213	37631.435080	13030.922551	26
18	E05010903	Hillfields	51.473702	-2.522575	38159.595960	9266.134680	27

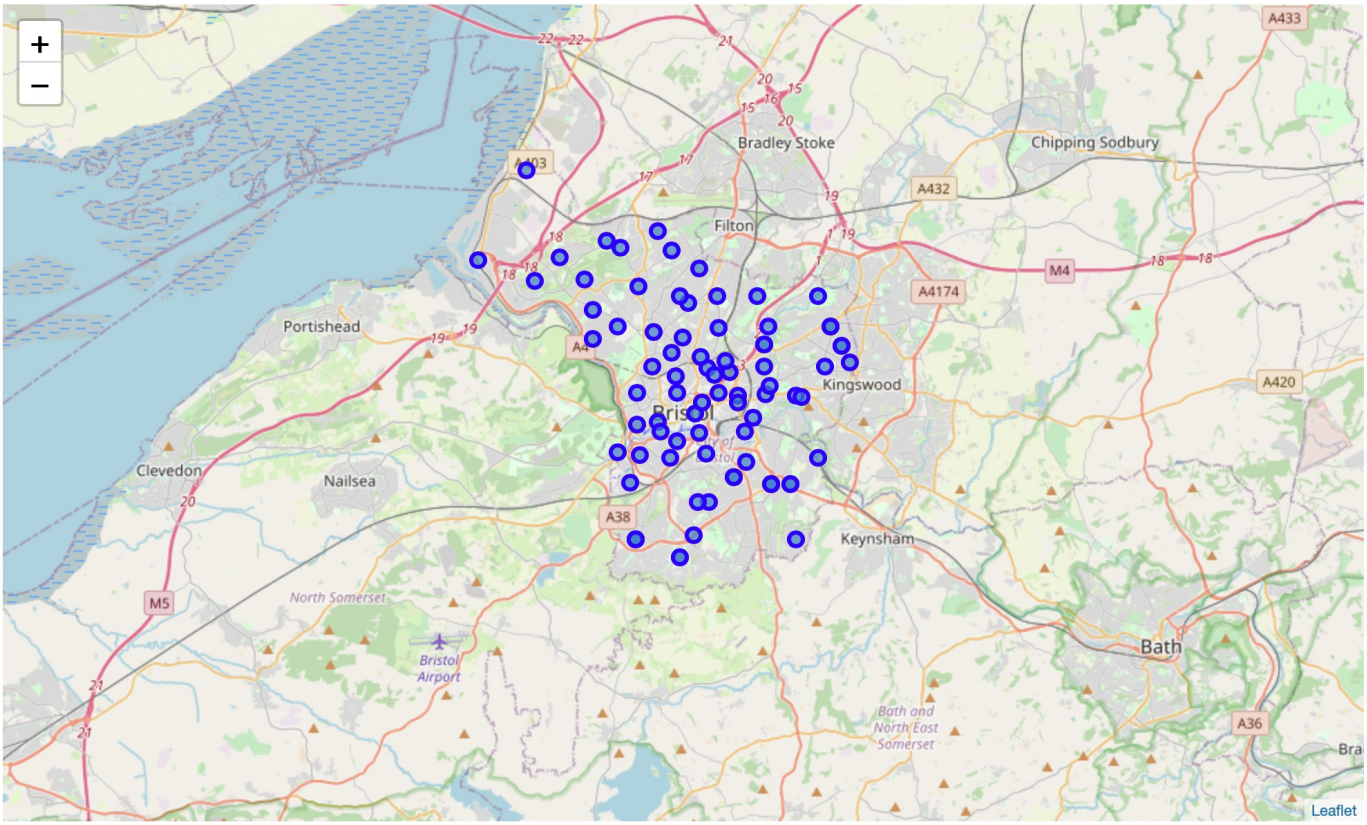
	Ward Code	Ward Name	Ward Latitude	Ward Longitude	Average Income	Index of Multiple Deprivation	A
19	E05010904	Horfield	51.495084	-2.586379	43355.555556	14102.841270	31
20	E05010905	Hotwells & Harbourside	51.449729	-2.609579	47079.695431	16354.888325	39
21	E05010906	Knowle	51.432777	-2.572334	40662.015504	12874.073643	30
22	E05010907	Lawrence Hill	51.454593	-2.571234	33904.667864	3005.393178	22
23	E05010908	Lockleaze	51.485856	-2.565667	38729.113924	7609.586498	32
24	E05010909	Redland	51.475245	-2.601158	56299.016393	26272.042623	51
25	E05010910	St George Central	51.462141	-2.526383	36888.888889	9147.989583	28
26	E05010911	St George Troopers Hill	51.453436	-2.530323	43668.072289	17611.259036	23
27	E05010912	St George West	51.460117	-2.546940	38075.294118	9440.205882	23
28	E05010913	Southmead	51.503062	-2.601118	36750.561798	5976.393258	29
29	E05010914	Southville	51.443322	-2.603991	43858.620690	12797.871473	34
30	E05010915	Stockwood	51.418419	-2.549735	37129.880478	13349.601594	22
31	E05010916	Stoke Bishop	51.477426	-2.636043	52207.272727	23625.712121	59
32	E05010917	Westbury-on-Trym & Henleaze	51.489795	-2.617483	54507.500000	29341.921429	53
33	E05010918	Windmill Hill	51.439097	-2.585944	45177.238806	13710.276119	30

One of task of creating dataset was to have a mapping between neighborhood – ward and LSOA.

The Lsoa is interesting to use because can identify in the map a punctual area and gives well recognizable detail of it¶

	name	latitude	longitude	Ward Code	Ward Name	Ward Latitude	Ward Longitude	LSOA11 Local
0	City centre	51.453632	-2.591341	E05010892	Central	51.454413	-2.594338	Redcliffe North
1	Arnos Vale	51.431600	-2.553600	E05010891	Brislington West	51.434392	-2.549071	Callington Road
2	Ashley Down	51.480000	-2.580000	E05010888	Bishopston & Ashley Down	51.481452	-2.586465	Ashley Down
3	Ashton Gate	51.440600	-2.619050	E05010887	Bedminster	51.437159	-2.621875	Bower Ashton
4	Ashton Vale	51.432193	-2.623522	E05010887	Bedminster	51.437159	-2.621875	Ashton Vale
...	...	...	...	...	...	...	...	...
76	Westbury Park	51.478800	-2.612100	E05010917	Westbury-on-Trym & Henleaze	51.489794	-2.617485	North View
77	Whitehall	51.462000	-2.554000	E05010912	St George West	51.460116	-2.546940	Redfield
78	Windmill Hill	51.441200	-2.586100	E05010918	Windmill Hill	51.439096	-2.585946	Victoria Park
79	Withywood	51.414770	-2.621140	E05010889	Bishopsworth	51.421787	-2.616458	Kings Head Par
83	Conham	51.440000	-2.530000	E05010890	Brislington East	51.440282	-2.538654	Broomhill Road

The dataset is: 81 rows — 9 columns



Fetching the restaurants through Foursquare API¶

A this point of time the exploration of 1 km radius for each of each neighbourhoods in order to lookup the Restaurant with price equals to **3** and **4**, where the assumption is: *the expensive restaurant* has good quality cuisine.

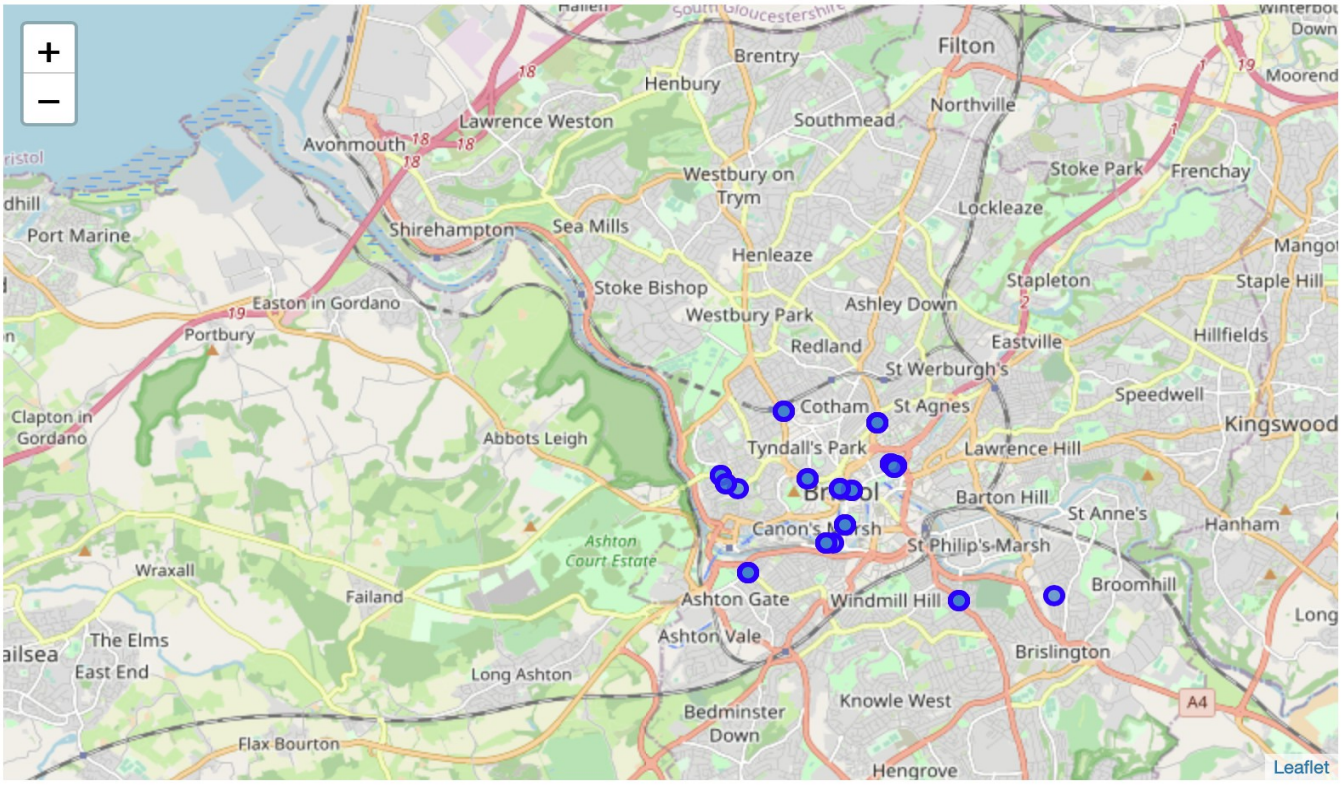
Here the first 5 rows of 69 records:

	name	latitude	longitude	Venue	Venue Id	Venue Latitude	Venue Longitud
3	City centre	51.453632	-2.591341	CÃ´te Brasserie	588c7f920319b8631e76e17a	51.457400	-2.587210
4	City centre	51.453632	-2.591341	Harvey Nichols Restaurant & Bar	50450455183fcb060e8d47e7	51.457233	-2.586228
5	City centre	51.453632	-2.591341	Pata Negra	5409ee3f498ea27289eb00e1	51.453950	-2.595134
10	City centre	51.453632	-2.591341	El Puerto	4e8748f8d3e393c4a795ad5d	51.449485	-2.596547
11	City centre	51.453632	-2.591341	The Olive Shed	4b05883df964a520fab922e3	51.447087	-2.600167

Interesting to count how many restaurants there are for each neighboorhood.

	name	latitude	longitude	count
0	Ashton Gate	51.440600	-2.619050	1
1	Bedminster	51.439700	-2.603700	3
2	Bower Ashton	51.441660	-2.629880	1
3	Broadmead	51.457000	-2.588000	6
4	City centre	51.453632	-2.591341	8
5	Clifton	51.460000	-2.620000	3
6	Cliftonwood	51.451098	-2.609712	7

	name	latitude	longitude	count
7	Cotham	51.465300	-2.601200	2
8	Hotwells	51.450000	-2.620000	4
9	Kingsdown	51.468000	-2.613000	1
10	Knowle	51.434000	-2.572100	1
11	Montpelier	51.467867	-2.585488	1
12	Redcliffe	51.447800	-2.589500	4
13	Redland	51.472410	-2.603210	1
14	Southville	51.445000	-2.600300	2
15	Spike Island	51.447900	-2.608700	4
16	St Andrew's	51.471280	-2.588772	1
17	St. Jude's	51.460012	-2.579625	4
18	St. Paul's	51.465600	-2.582000	4
19	Totterdown	51.438800	-2.566000	2
20	Tyndall's Park	51.460000	-2.600000	7
21	Upper Knowle	51.434000	-2.572100	1
22	Windmill Hill	51.441200	-2.586100	1



Only the following neighborhoods have restaurants in nearby:

```
['Ashton Gate' 'Bedminster' 'Bower Ashton' 'Broadmead' 'City centre'
 'Clifton' 'Cliftonwood' 'Cotham' 'Hotwells' 'Kingsdown' 'Knowle'
 'Montpelier' 'Redcliffe' 'Redland' 'Southville' 'Spike Island'
```

```
"St Andrew's" "St. Jude's" "St. Paul's" 'Totterdown' "Tyndall's Park"
'Upper Knowle' 'Windmill Hill']
```

The venue dataset will be reviewed in order to create a column for each specific restaurant and calculate the mean of that cuisine for the specific neighborhood.

	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New Ameri
0	Ashton Gate	51.440600	-2.619050	0.0	0.00	0.0	0.00	0.00
1	Bedminster	51.439700	-2.603700	0.0	0.00	0.0	0.00	0.00
2	Bower Ashton	51.441660	-2.629880	0.0	0.00	0.0	0.00	0.00
3	Broadmead	51.457000	-2.588000	0.0	0.33	0.0	0.17	0.17
4	City centre	51.453632	-2.591341	0.0	0.38	0.0	0.12	0.12
5	Clifton	51.460000	-2.620000	0.0	0.33	0.0	0.00	0.00
6	Cliftonwood	51.451098	-2.609712	0.0	0.29	0.0	0.00	0.14
7	Cotham	51.465300	-2.601200	0.0	0.00	0.0	0.00	0.00
8	Hotwells	51.450000	-2.620000	0.0	0.25	0.0	0.00	0.00
9	Kingsdown	51.468000	-2.613000	0.0	0.00	0.0	0.00	0.00
10	Knowle	51.434000	-2.572100	0.0	0.00	1.0	0.00	0.00
11	Montpelier	51.467867	-2.585488	0.0	0.00	0.0	0.00	0.00
12	Redcliffe	51.447800	-2.589500	0.0	0.00	0.0	0.00	0.25
13	Redland	51.472410	-2.603210	0.0	0.00	0.0	0.00	0.00
14	Southville	51.445000	-2.600300	0.0	0.00	0.0	0.00	0.00
15	Spike Island	51.447900	-2.608700	0.0	0.25	0.0	0.00	0.00
16	St Andrew's	51.471280	-2.588772	0.0	0.00	0.0	0.00	0.00
17	St. Jude's	51.460012	-2.579625	0.0	0.50	0.0	0.25	0.00
18	St. Paul's	51.465600	-2.582000	0.0	0.50	0.0	0.25	0.00
19	Totterdown	51.438800	-2.566000	0.5	0.00	0.5	0.00	0.00
20	Tyndall's Park	51.460000	-2.600000	0.0	0.29	0.0	0.14	0.14
21	Upper Knowle	51.434000	-2.572100	0.0	0.00	1.0	0.00	0.00
22	Windmill Hill	51.441200	-2.586100	0.0	0.00	1.0	0.00	0.00

Adding crime - house price - income - population - deprivation index¶

For each of five information we are going to create specific ranges based on 5 bins.

	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	count rest. nationality	War Nam
0	Ashton Gate	51.440600	-2.619050	0.0	0.00	0.0	0.00	0.00	1.00	1	Bedmin
1	Bedminster	51.439700	-2.603700	0.0	0.00	0.0	0.00	0.00	1.00	1	Bedmin
2	Bower	51.441660	-2.629880	0.0	0.00	0.0	0.00	0.00	1.00	1	Bedmin

	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	count rest. nationality	War Nam
	Ashton										
3	Broadmead	51.457000	-2.588000	0.0	0.33	0.0	0.17	0.17	0.33	4	Central
4	City centre	51.453632	-2.591341	0.0	0.38	0.0	0.12	0.12	0.38	4	Central

Divide in bins¶

In order simplify the characteristics, the cut in bins has been applied and they are the following meaning:

- Diversity Restaurant:
    - [1] very few, [2] few, [3] medium, [4] Many, [5] a lot
- Population:
    - [1] very low, [2] low, [3] medium, [4] high, [5] very high
- Crimes:
    - [5] very low, [4] low, [3] medium, [2] bad, [1] very bad
- Avg. House Price £k:
    - [1] very low", [2] low, [3] medium, [4] high, [5] very high
- Average Income:
    - [1] very low", [2] low, [3] medium, [4] high, [5] very high
- Deprivation Index:
    - [1] very low", [2] low, [3] medium, [4] high, [5] very high

	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restauran Ranges
0	Ashton Gate	51.440600	-2.619050	0.0	0.00	0.0	0.00	0.00	1.00	1
1	Bedminster	51.439700	-2.603700	0.0	0.00	0.0	0.00	0.00	1.00	1
2	Bower Ashton	51.441660	-2.629880	0.0	0.00	0.0	0.00	0.00	1.00	1
3	Broadmead	51.457000	-2.588000	0.0	0.33	0.0	0.17	0.17	0.33	5
4	City centre	51.453632	-2.591341	0.0	0.38	0.0	0.12	0.12	0.38	5
5	Clifton	51.460000	-2.620000	0.0	0.33	0.0	0.00	0.00	0.67	2
6	Cliftonwood	51.451098	-2.609712	0.0	0.29	0.0	0.00	0.14	0.57	4
7	Cotham	51.465300	-2.601200	0.0	0.00	0.0	0.00	0.00	1.00	1
8	Hotwells	51.450000	-2.620000	0.0	0.25	0.0	0.00	0.00	0.75	2
9	Kingsdown	51.468000	-2.613000	0.0	0.00	0.0	0.00	0.00	1.00	1
10	Knowle	51.434000	-2.572100	0.0	0.00	1.0	0.00	0.00	0.00	1
11	Montpelier	51.467867	-2.585488	0.0	0.00	0.0	0.00	0.00	1.00	1
12	Redcliffe	51.447800	-2.589500	0.0	0.00	0.0	0.00	0.25	0.75	2
13	Redland	51.472410	-2.603210	0.0	0.00	0.0	0.00	0.00	1.00	1

	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restauran Ranges
14	Southville	51.445000	-2.600300	0.0	0.00	0.0	0.00	0.00	1.00	1
15	Spike Island	51.447900	-2.608700	0.0	0.25	0.0	0.00	0.00	0.75	2
16	St Andrew's	51.471280	-2.588772	0.0	0.00	0.0	0.00	0.00	1.00	1
17	St. Jude's	51.460012	-2.579625	0.0	0.50	0.0	0.25	0.00	0.25	4
18	St. Paul's	51.465600	-2.582000	0.0	0.50	0.0	0.25	0.00	0.25	4
19	Totterdown	51.438800	-2.566000	0.5	0.00	0.5	0.00	0.00	0.00	2
20	Tyndall's Park	51.460000	-2.600000	0.0	0.29	0.0	0.14	0.14	0.43	5
21	Upper Knowle	51.434000	-2.572100	0.0	0.00	1.0	0.00	0.00	0.00	1
22	Windmill Hill	51.441200	-2.586100	0.0	0.00	1.0	0.00	0.00	0.00	1

Methodology ¶

Now we are going to clusterize the localities in order to identify the most suitable for the stackholder

In order to find a potential cluster k, the Clusterizing process will be first looped and check a value of inertia as trade of.

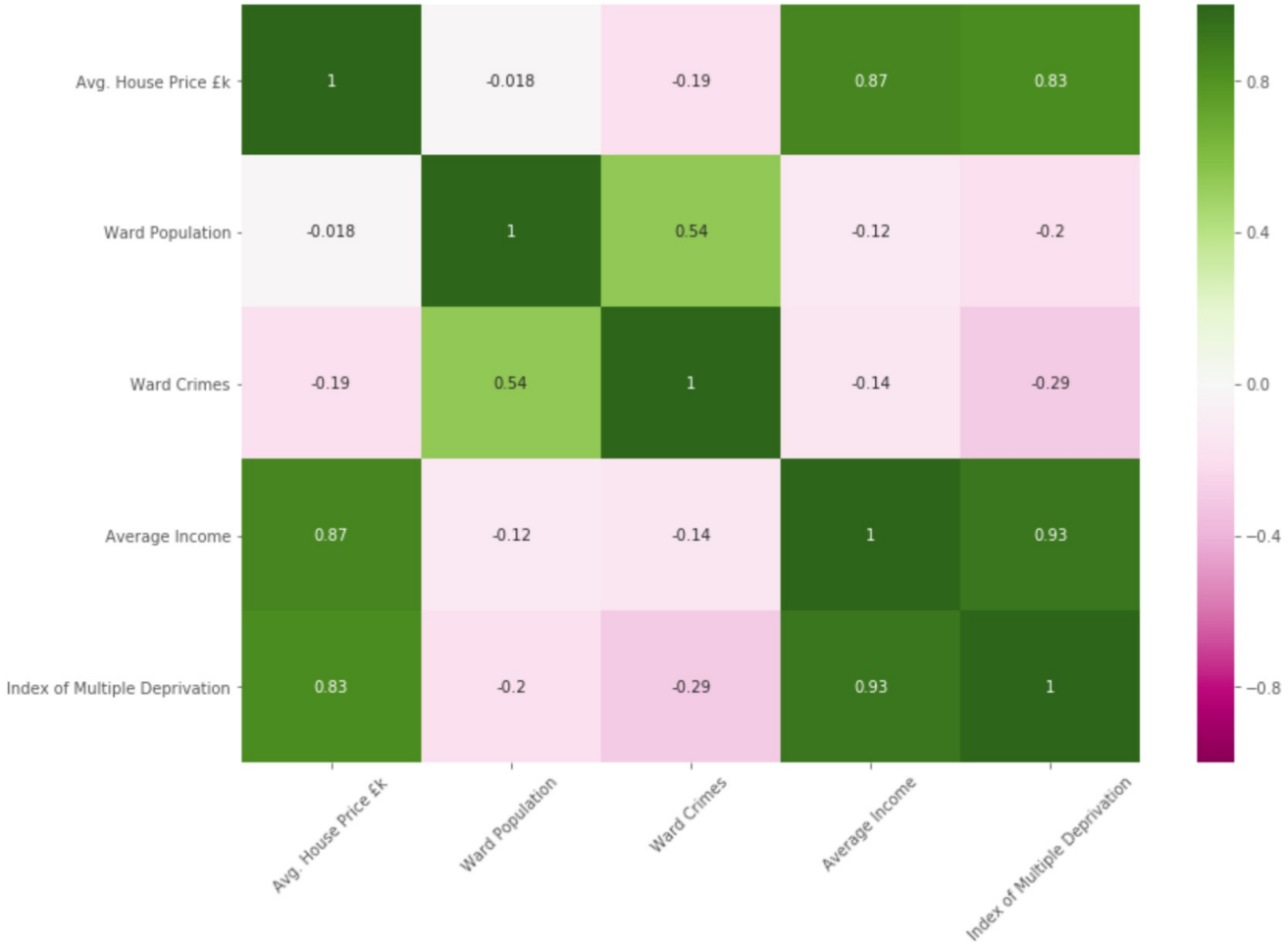
Analysis ¶

Ward Correlation analysis¶

The following section will show the correlation between 'Avg. House Price Â£k', 'Ward Population', 'Ward Crimes', 'Average Income', 'Index of Multiple Deprivation'.

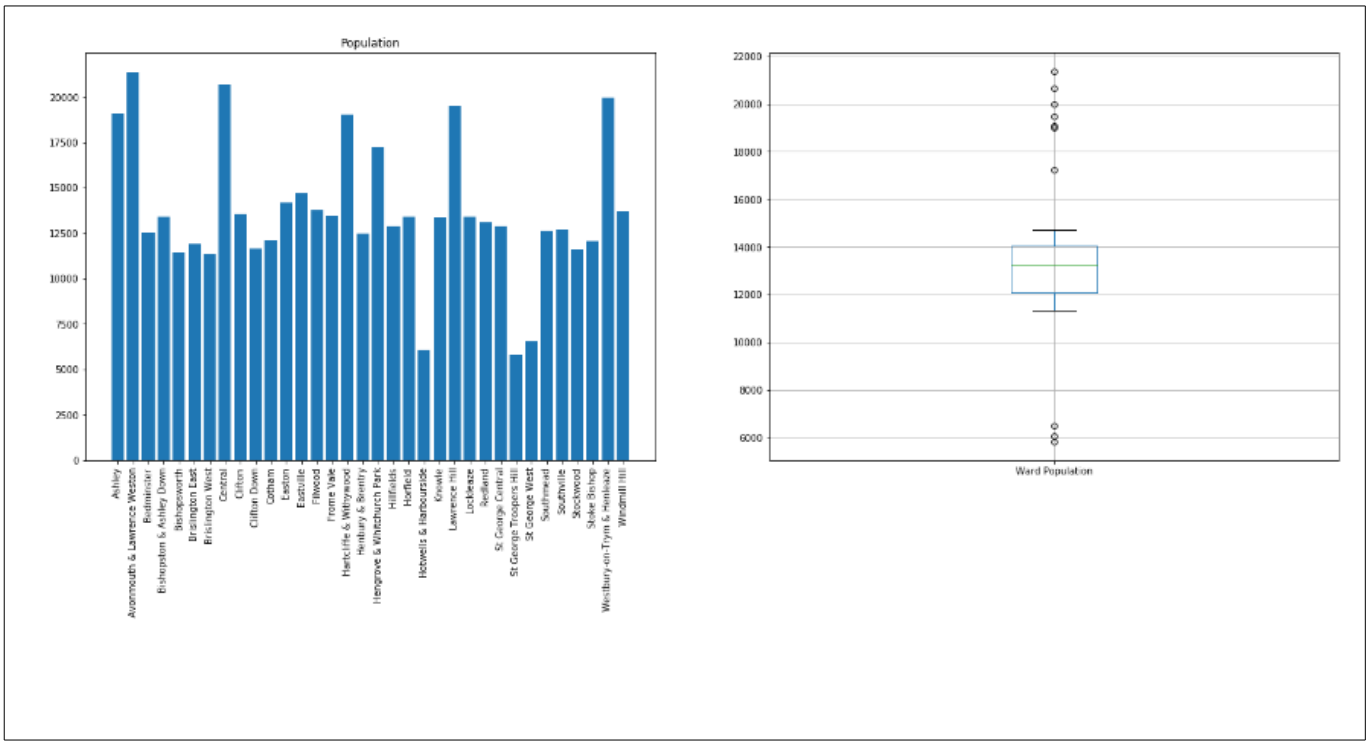
	Avg. House Price Â£k	Ward Population	Ward Crimes	Average Income	Index of Multiple Deprivatio
Avg. House Price Â£k	1.000000	-0.017770	-0.189690	0.865146	0.832418
Ward Population	-0.017770	1.000000	0.543593	-0.118037	-0.195242
Ward Crimes	-0.189690	0.543593	1.000000	-0.144232	-0.292878
Average Income	0.865146	-0.118037	-0.144232	1.000000	0.927073
Index of Multiple Deprivation	0.832418	-0.195242	-0.292878	0.927073	1.000000

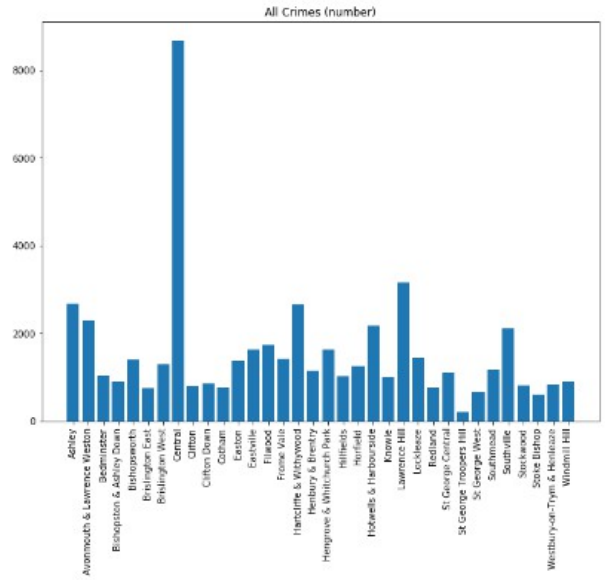
```
Avg. House Price Â£k             float64
Ward Population                 float64
Ward Crimes                     float64
Average Income                  float64
Index of Multiple Deprivation   float64
dtype: object
```

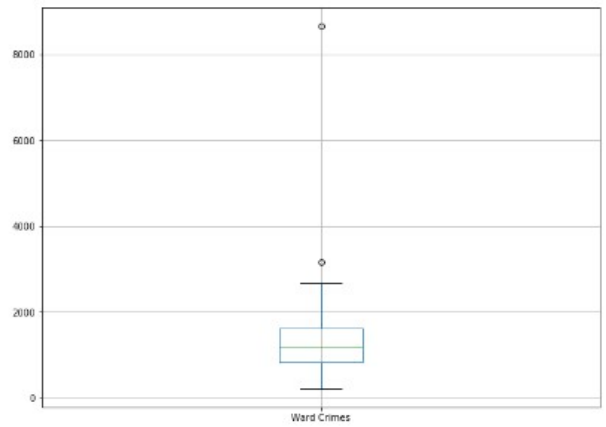


As show there are strong positive correlations between *Index of Multiple Deprivation*, *Average Income* and *Avg. House Price £k*.

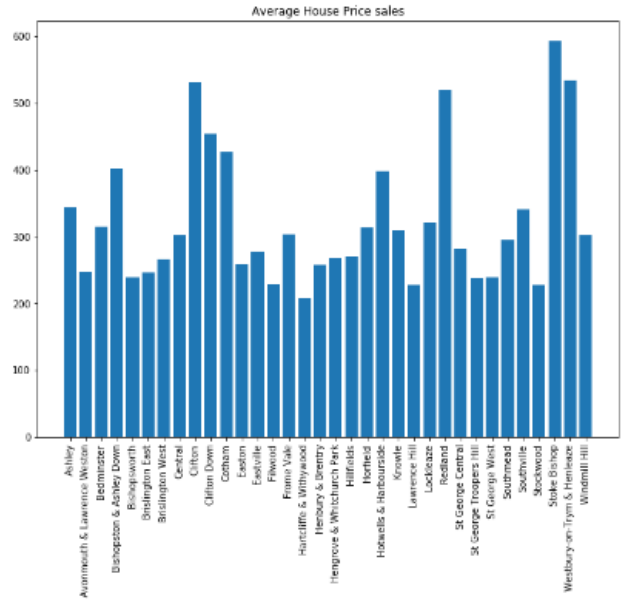
Additionally the following bar plots can show those wards can be over the average or a specific value.

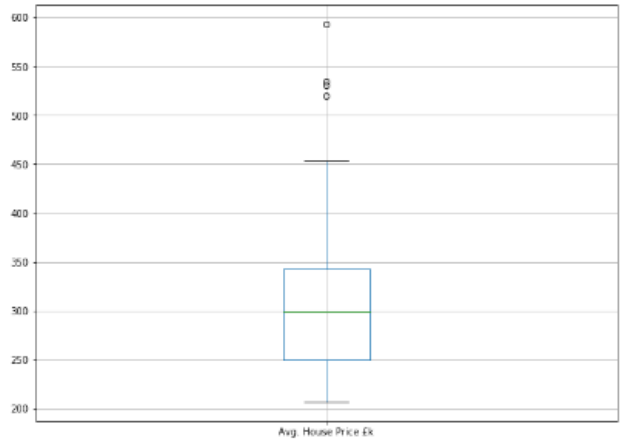




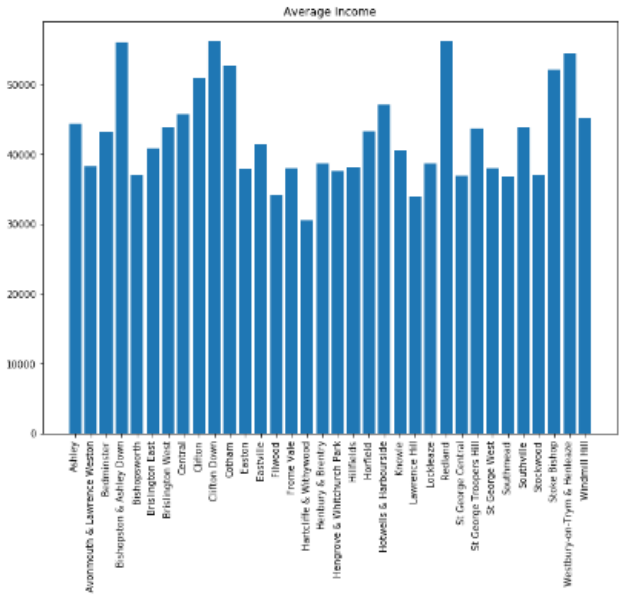


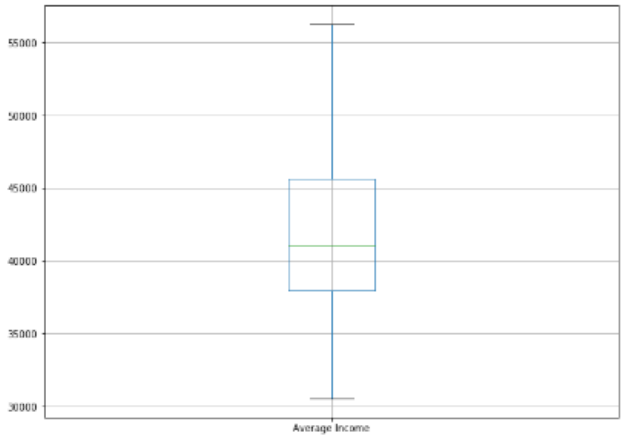
About the crime, the Central city is the outlier with biggest value.



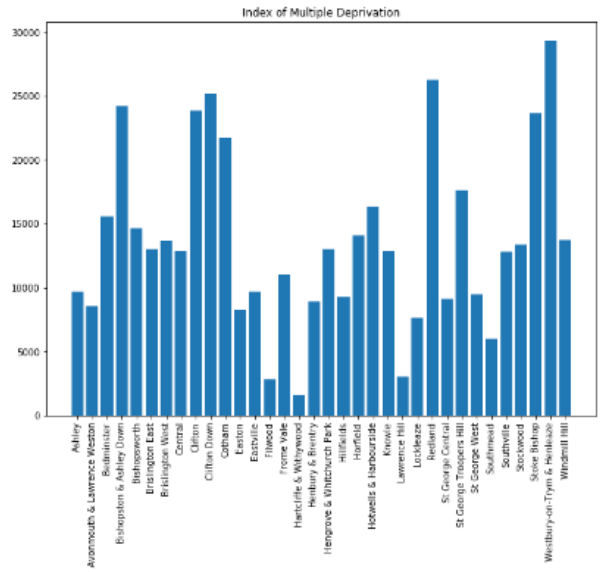


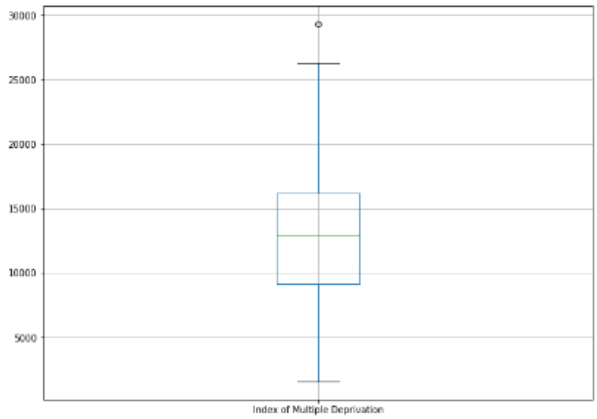
About the Average House Price sales, the price is around 300k£, the outliers





The average income is 42k per person, the high values are in Clifton, Redland and Ashley





The areas with lower deprivation are Clifton, Westbury-on-Trym, Redland.

-- Ward Crimes --

	Ward Code	Ward Name	Ward Crimes
26	E05010911	St George Troopers Hill	209.0
31	E05010916	Stoke Bishop	604.0
27	E05010912	St George West	669.0
5	E05010890	Brislington East	751.0
24	E05010909	Redland	772.0

-- Average Income --

	Ward Code	Ward Name	Average Income
24	E05010909	Redland	56299.016393
9	E05010894	Clifton Down	56292.763158
3	E05010888	Bishopston & Ashley Down	56094.258373
32	E05010917	Westbury-on-Trym & Henleaze	54507.500000
10	E05010895	Cotham	52680.487805

-- Population --

	Ward Code	Ward Name	Ward Population
1	E05010886	Avonmouth & Lawrence Weston	21375.0
7	E05010892	Central	20656.0
32	E05010917	Westbury-on-Trym & Henleaze	19971.0
22	E05010907	Lawrence Hill	19473.0
0	E05010885	Ashley	19083.0

-- Avg. House Price Â£k --

	Ward Code	Ward Name	Avg. House Price Â£k
31	E05010916	Stoke Bishop	593.612121
32	E05010917	Westbury-on-Trym & Henleaze	534.566071
8	E05010893	Clifton	530.602273
24	E05010909	Redland	519.613115
9	E05010894	Clifton Down	453.302632

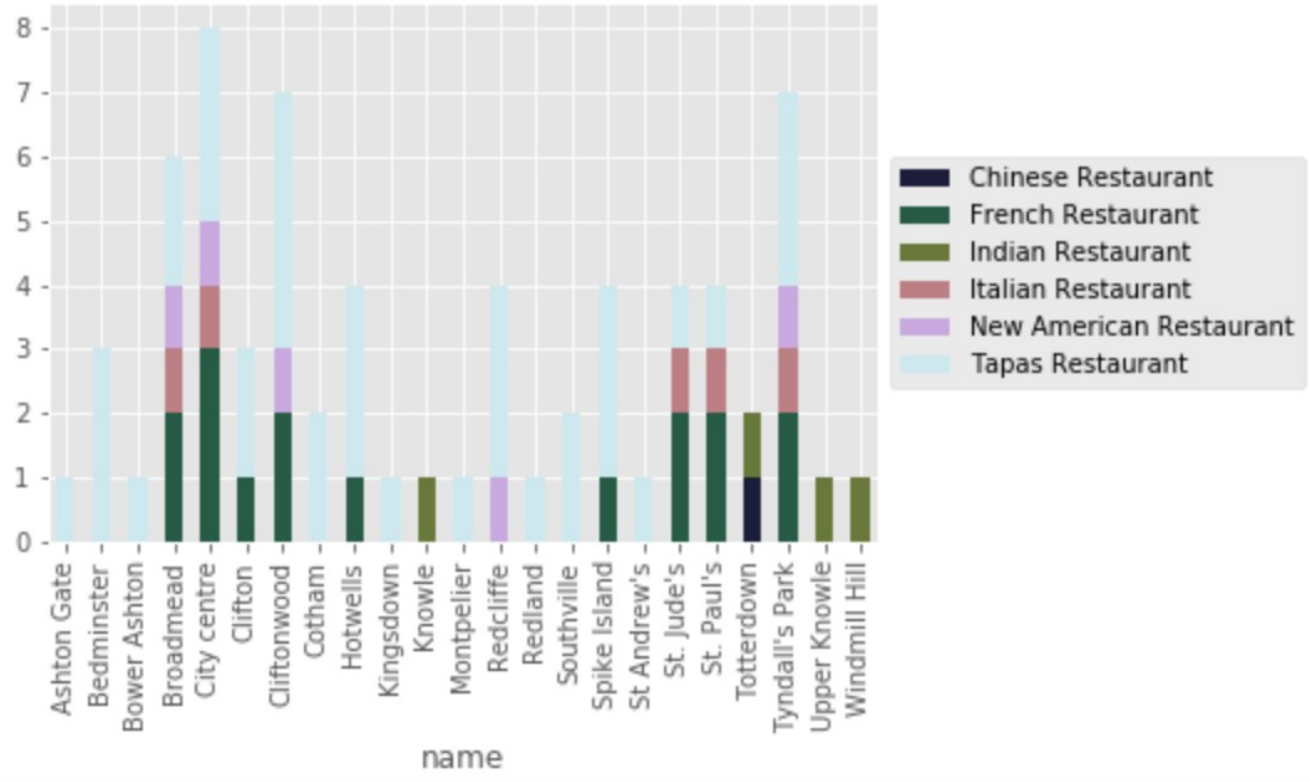
```
-- Index of Multiple Deprivation --
    Ward Code                    Ward Name  Index of Multiple Deprivation
32  E05010917  Westbury-on-Trym & Henleaze                   29341.921429
24  E05010909                     Redland                   26272.042623
9   E05010894                 Clifton Down                   25230.625000
3   E05010888     Bishopston & Ashley Down                   24239.947368
8   E05010893                      Clifton                   23870.107955
```

Locality venues¶

The following diagrams will show the number of venues close to the locality.

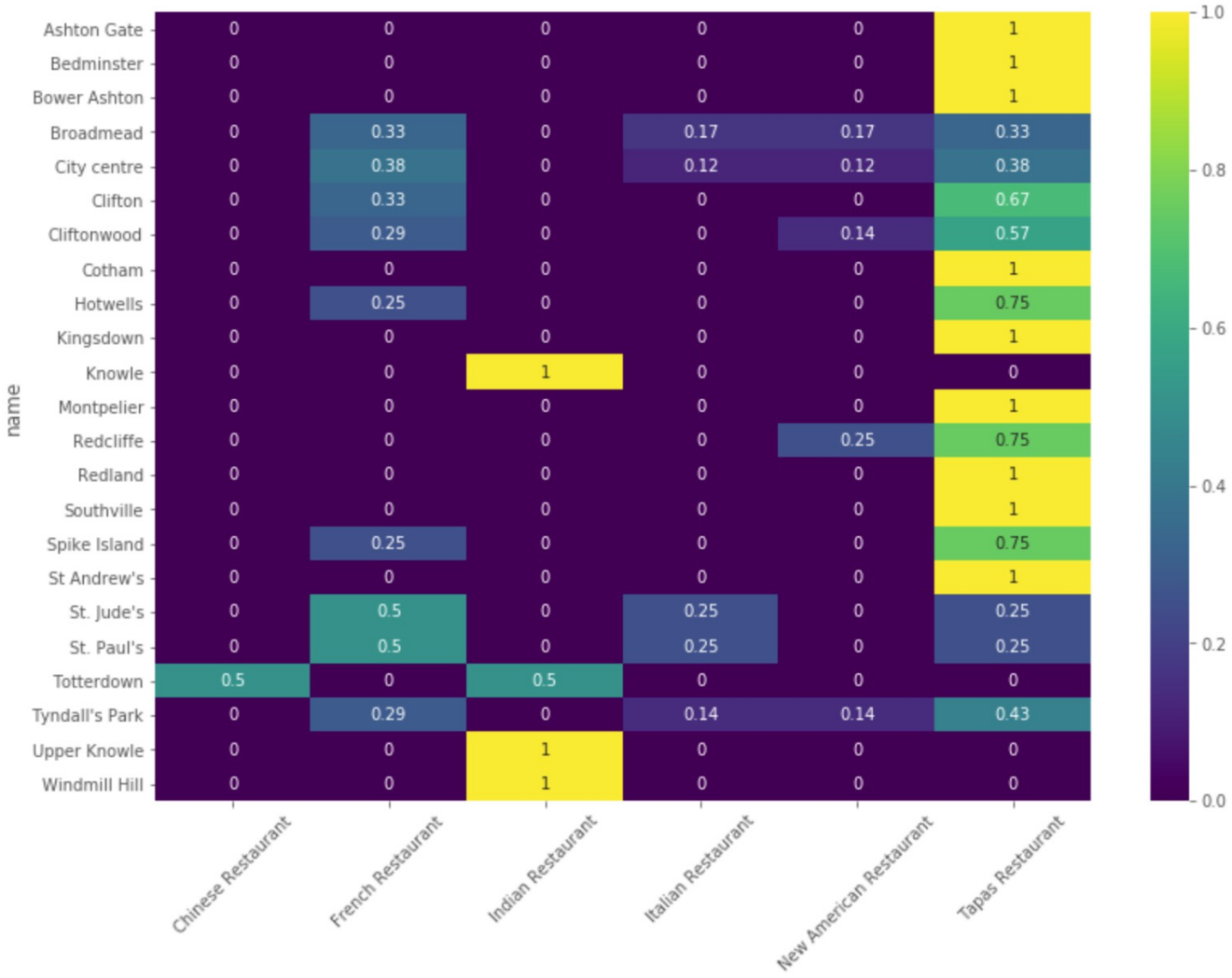
	name	latitude	longitude	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New Ameri
3	Broadmead	51.457000	-2.588000	0.0	0.33	0.0	0.17	0.17
4	City centre	51.453632	-2.591341	0.0	0.38	0.0	0.12	0.12
20	Tyndall's Park	51.460000	-2.600000	0.0	0.29	0.0	0.14	0.14
6	Cliftonwood	51.451098	-2.609712	0.0	0.29	0.0	0.00	0.14
18	St. Paul's	51.465600	-2.582000	0.0	0.50	0.0	0.25	0.00

```
Available restaurant for neighborhood:

['Chinese Restaurant', 'French Restaurant', 'Indian Restaurant',
       'Italian Restaurant', 'New American Restaurant', 'Tapas Restaurant']
```



Heatmap¶

The **heatmap** helps to identify which kind of restaturant is located close to the location and shows the distribution

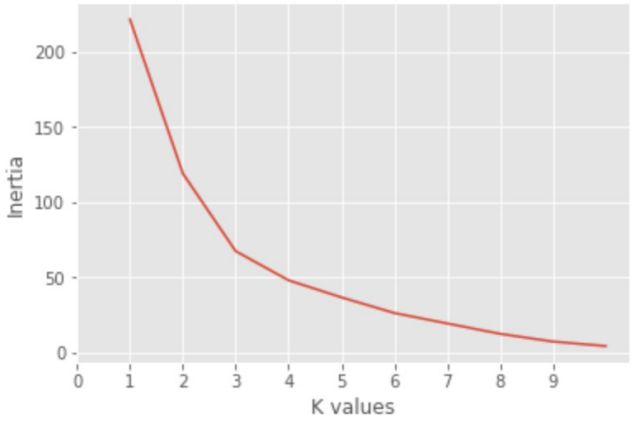


Clustering¶

The algorithm of clustering will gather those neighbourhoods with close characteristics.

But in order to identify a trade off value of cluster we are going to analyze the *inertia* for the *k* from *1* to *10*

```
k: 1  cost: 220.26483240484234
k: 2  cost: 119.11713382136685
k: 3  cost: 67.75391551037788
k: 4  cost: 48.31226525585979
k: 5  cost: 36.937179545354
k: 6  cost: 26.973847367491498
k: 7  cost: 20.012398901785946
k: 8  cost: 13.116907664841282
k: 9  cost: 8.236765754548031
k: 10  cost: 5.236747321363531
```

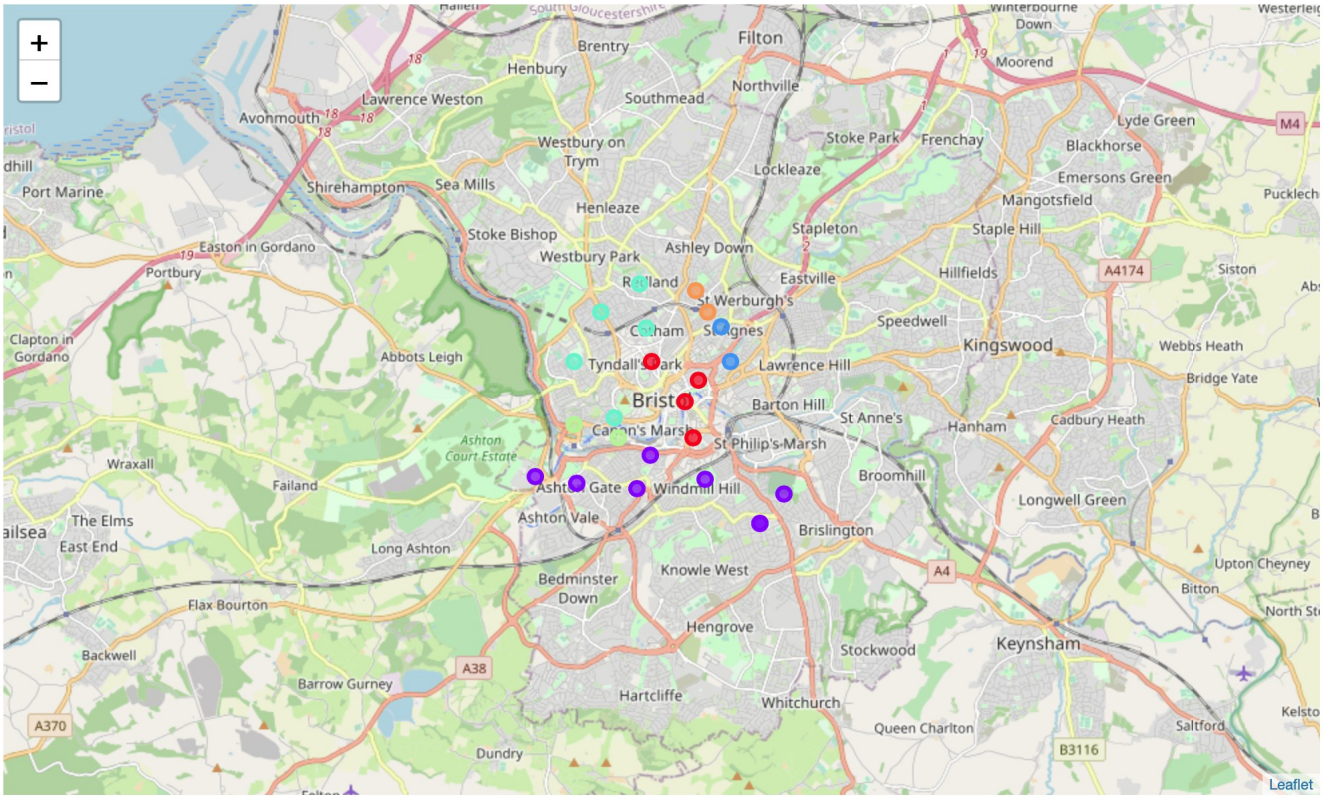


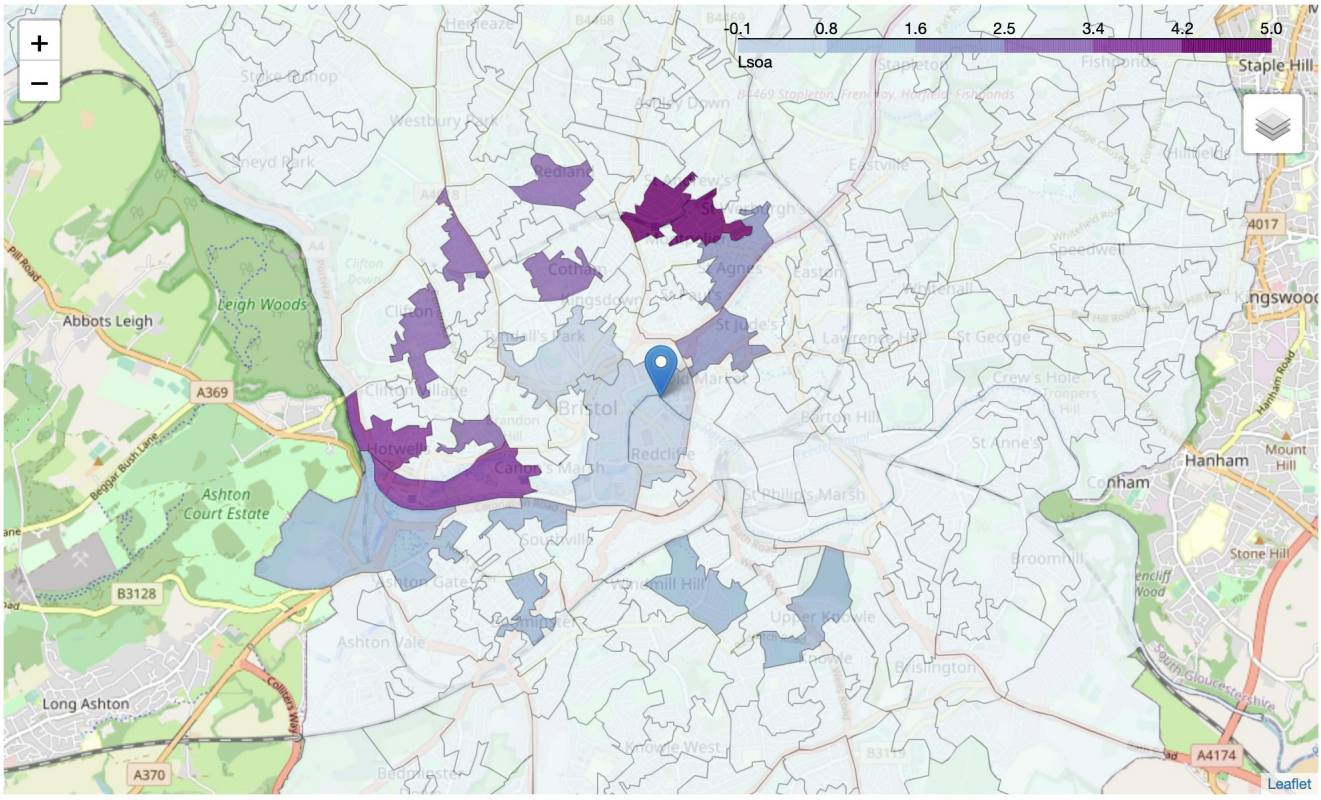
By plotting the k's a valid value is **6**

Defined k = 6 then we define the single clusters

	Cluster Labels	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Population Ranges	Crim Rang
0	1	Ashton Gate	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
1	1	Bedminster	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
2	1	Bower Ashton	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
3	0	Broadmead	0.0	0.33	0.0	0.17	0.17	0.33	5	5	1
4	0	City centre	0.0	0.38	0.0	0.12	0.12	0.38	5	5	1
5	3	Clifton	0.0	0.33	0.0	0.00	0.00	0.67	2	3	5
6	3	Cliftonwood	0.0	0.29	0.0	0.00	0.14	0.57	4	3	5
7	3	Cotham	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
8	4	Hotwells	0.0	0.25	0.0	0.00	0.00	0.75	2	1	5
9	3	Kingsdown	0.0	0.00	0.0	0.00	0.00	1.00	1	2	5
10	1	Knowle	0.0	0.00	1.0	0.00	0.00	0.00	1	3	5
11	5	Montpelier	0.0	0.00	0.0	0.00	0.00	1.00	1	5	4
12	0	Redcliffe	0.0	0.00	0.0	0.00	0.25	0.75	2	5	1
13	3	Redland	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
14	1	Southville	0.0	0.00	0.0	0.00	0.00	1.00	1	3	5
15	4	Spike Island	0.0	0.25	0.0	0.00	0.00	0.75	2	1	5
16	5	St Andrew's	0.0	0.00	0.0	0.00	0.00	1.00	1	5	4
17	2	St. Jude's	0.0	0.50	0.0	0.25	0.00	0.25	4	5	4
18	2	St. Paul's	0.0	0.50	0.0	0.25	0.00	0.25	4	5	4

	Cluster Labels	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Population Ranges	Crim Rang
19	1	Totterdown	0.5	0.00	0.5	0.00	0.00	0.00	2	3	5
20	0	Tyndall's Park	0.0	0.29	0.0	0.14	0.14	0.43	5	5	1
21	1	Upper Knowle	0.0	0.00	1.0	0.00	0.00	0.00	1	3	5
22	1	Windmill Hill	0.0	0.00	1.0	0.00	0.00	0.00	1	3	5





	Cluster Labels	name	Ward Name	LSOA11 Local name
0	1	Ashton Gate	Bedminster	Bower Ashton
1	1	Bedminster	Bedminster	West Street West
2	1	Bower Ashton	Bedminster	Bower Ashton
3	0	Broadmead	Central	City Centre and Queens Square
4	0	City centre	Central	Redcliffe North
5	3	Clifton	Clifton	Clifton College
6	3	Cliftonwood	Clifton	Cliftonwood
7	3	Cotham	Cotham	Cotham Park
8	4	Hotwells	Hotwells & Harbourside	Hotwells
9	3	Kingsdown	Clifton Down	Whatley Road
10	1	Knowle	Knowle	Upper Knowle
11	5	Montpelier	Ashley	Upper Montpelier
12	0	Redcliffe	Central	Redcliffe North
13	3	Redland	Redland	Redland Court Road
14	1	Southville	Southville	Coronation Road East
15	4	Spike Island	Hotwells & Harbourside	Spike Island
16	5	St Andrew's	Ashley	Cromwell Road
17	2	St. Jude's	Lawrence Hill	Cabot Circus
18	2	St. Paul's	Ashley	St Agnes
19	1	Totterdown	Knowle	Upper Knowle

	Cluster Labels	name	Ward Name	LSOA11 Local name
20	0	Tyndall's Park	Central	University
21	1	Upper Knowle	Knowle	Upper Knowle
22	1	Windmill Hill	Windmill Hill	Victoria Park

Cluster 1¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Popula Range
3	Broadmead	0.0	0.33	0.0	0.17	0.17	0.33	5	5
4	City centre	0.0	0.38	0.0	0.12	0.12	0.38	5	5
12	Redcliffe	0.0	0.00	0.0	0.00	0.25	0.75	2	5
20	Tyndall's Park	0.0	0.29	0.0	0.14	0.14	0.43	5	5

One of characteristic of the cluster 1 is to gather neighbourhood with high population but with high crimes and most of them have highly diversity cuisine.

Cluster 2¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Popula Range
0	Ashton Gate	0.0	0.0	0.0	0.0	0.0	1.0	1	3
1	Bedminster	0.0	0.0	0.0	0.0	0.0	1.0	1	3
2	Bower Ashton	0.0	0.0	0.0	0.0	0.0	1.0	1	3
10	Knowle	0.0	0.0	1.0	0.0	0.0	0.0	1	3
14	Southville	0.0	0.0	0.0	0.0	0.0	1.0	1	3
19	Totterdown	0.5	0.0	0.5	0.0	0.0	0.0	2	3
21	Upper Knowle	0.0	0.0	1.0	0.0	0.0	0.0	1	3
22	Windmill Hill	0.0	0.0	1.0	0.0	0.0	0.0	1	3

The cluster 2 has a bad diversity of restaurant but the crimes are very low.

Cluster 3¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Populatio Ranges
17	St.	0.0	0.5	0.0	0.25	0.0	0.25	4	5

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Populatio Ranges
	Jude's								
18	St. Paul's	0.0	0.5	0.0	0.25	0.0	0.25	4	5

The cluster 3 has neighbourhoods highly populated and the deprivation is high.

Cluster 4¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Populat Range
5	Clifton	0.0	0.33	0.0	0.0	0.00	0.67	2	3
6	Cliftonwood	0.0	0.29	0.0	0.0	0.14	0.57	4	3
7	Cotham	0.0	0.00	0.0	0.0	0.00	1.00	1	3
9	Kingsdown	0.0	0.00	0.0	0.0	0.00	1.00	1	2
13	Redland	0.0	0.00	0.0	0.0	0.00	1.00	1	3

The cluster 4 highlights neighbourhoods quite rich and the quality of live is high: low crime, very good income, low deprivation, the population is in the average, but only one of them it's interesting *Cliftonwood* because the only one with high restaurant diversity and additional there is not competitor for our Italian Chef.

Cluster 5¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Populati Range
8	Hotwells	0.0	0.25	0.0	0.0	0.0	0.75	2	1
15	Spike Island	0.0	0.25	0.0	0.0	0.0	0.75	2	1

The cluster 5 is highly populated and house / income / deprivation are in the average, low crime but only few diversity restaurant.

Cluster 6¶

	name	Chinese Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Diversity Restaurant Ranges	Populat Range
11	Montpelier	0.0	0.0	0.0	0.0	0.0	1.0	1	5
16	St Andrew's	0.0	0.0	0.0	0.0	0.0	1.0	1	5

The cluster 6 has basically only one type of cuisine *Tapas Restaurant*, the population is low, good in crimes and the average for the

rest of variables.

Results and Discussion ¶

After fetching datasets from several sources in different ways, downloading files, sgrabbing website and adopted method of cleaning, aggregation and merging, the final dataset of neighbourhoods has been sliced in 6 clusters.

Conclusion ¶

The **cluster 4** shows in *Cliftonwood* the area which is very close to the needs of stackholder.